IFW

PTO/SB/21 (09-04)

TRANSMITTAL FORM

(to be used for all correspondence after initial filing)

Total Number of Pages in This Submission	1

Application Number	10/754,164
Filing Date	January 9, 2004
First Named Inventor	NAKAGAWA, Yoshihito
Art Unit	2171
Examiner Name	Unassigned
Attorney Docket Number	16869P-102400US

ENCLOSURES *(Check all that apply)*

- ☐ Fee Transmittal Form
  - ☐ Fee Attached
- ☐ Amendment/Reply
  - ☐ After Final
  - ☐ Affidavits/declaration(s)
- ☐ Extension of Time Request
- ☐ Express Abandonment Request
- ☐ Information Disclosure Statement
- ☒ Certified Copy of Priority Document(s)
- ☐ Reply to Missing Parts/ Incomplete Application
  - ☐ Reply to Missing Parts under 37 CFR 1.52 or 1.53
- ☐ Drawing(s)
- ☐ Licensing-related Papers
- ☐ Petition
- ☐ Petition to Convert to a Provisional Application
- ☐ Power of Attorney, Revocation Change of Correspondence Address
- ☐ Terminal Disclaimer
- ☐ Request for Refund
- ☐ CD, Number of CD(s) _____
  - ☐ Landscape Table on CD
- ☐ After Allowance Communication to TC
- ☐ Appeal Communication to Board of Appeals and Interferences
- ☐ Appeal Communication to TC **(Appeal Notice, Brief, Reply Brief)**
- ☐ Proprietary Information
- ☐ Status Letter
- ☒ Other Enclosure(s) (please identify below):
  Return Postcard

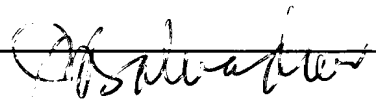
Remarks The Commissioner is authorized to charge any additional fees to Deposit Account 20-1430.

SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT

Firm Name	Townsend and Townsend and Crew LLP		
Signature			
Printed name	Chun-Pok Leung		
Date	December 10, 2004	Reg. No.	41,405

CERTIFICATE OF TRANSMISSION/MAILING

I hereby certify that this correspondence is being deposited with the United States Postal Service with sufficient postage as first class mail in an envelope addressed to: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450 on the date shown below.

Signature			
Typed or printed name	Joy Salvador	Date	December 10, 2004

日 本 国 特 許 庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出願年月日 Date of Application:	２００３年　６月　３日
出願番号 Application Number: [ST. 10/C]:	特願２００３－１５８２７１ [ＪＰ２００３－１５８２７１]
出願人 Applicant(s):	株式会社日立製作所

CERTIFIED COPY OF
PRIORITY DOCUMENT

２００４年　２月１９日

特許庁長官
Commissioner,
Japan Patent Office　今井康夫

出証番号　出証特２００４－３０１１０５８

【書類名】　特許願

【整理番号】　HI030203

【提出日】　平成15年 6月 3日

【あて先】　特許庁長官殿

【国際特許分類】　G06F　3/06

【発明者】

【住所又は居所】　神奈川県小田原市中里３２２番２号　株式会社日立製作所　ＲＡＩＤシステム事業部内

【氏名】　中川　義仁

【発明者】

【住所又は居所】　神奈川県小田原市中里３２２番２号　株式会社日立製作所　ＲＡＩＤシステム事業部内

【氏名】　小笠原　裕

【発明者】

【住所又は居所】　神奈川県小田原市中里３２２番２号　株式会社日立製作所　ＲＡＩＤシステム事業部内

【氏名】　内海　勝広

【発明者】

【住所又は居所】　神奈川県小田原市中里３２２番２号　株式会社日立製作所　ＲＡＩＤシステム事業部内

【氏名】　中山　信一

【発明者】

【住所又は居所】　神奈川県小田原市中里３２２番２号　株式会社日立製作所　ＲＡＩＤシステム事業部内

【氏名】　▲高▼田　豊

【特許出願人】

【識別番号】　000005108

【氏名又は名称】　株式会社日立製作所

【代理人】

【識別番号】　110000176

【氏名又は名称】　一色国際特許業務法人

【代表者】　一色　健輔

【手数料の表示】

【予納台帳番号】　211868

【納付金額】　21,000円

【提出物件の目録】

【物件名】　明細書　1

【物件名】　図面　1

【物件名】　要約書　1

【プルーフの要否】　要

【書類名】　明細書

【発明の名称】　ストレージ制御装置の制御方法及びストレージ制御装置

【特許請求の範囲】

【請求項１】　情報処理装置から送信されるデータ入出力要求を受信して、前記データ入出力要求に応じて記憶デバイスに対するデータの書き込み／読み出しを制御するための制御信号を出力する、互いに通信可能に接続される複数のチャネル制御部と、前記制御信号に応じて記憶デバイスに対するデータの書き込み／読み出しを行うディスク制御部と、を備えるストレージ制御装置の制御方法であって、

第１の前記チャネル制御部が、前記情報処理装置と通信することにより前記情報処理装置の動作を監視し、

第１の前記チャネル制御部が、前記情報処理装置における障害を検知した場合に、前記情報処理装置から送信されたデータ入出力要求が着信される第２の前記チャネル制御部により前記データ入出力要求に応じて実行される処理を制限するための処理を実行すること、

を特徴とするストレージ制御装置の制御方法。

【請求項２】　情報処理装置から送信されるデータ入出力要求を受信して前記データ入出力要求に応じて記憶デバイスに対するデータの書き込み／読み出しを制御するための制御信号を出力する、互いに通信可能に接続される複数のチャネル制御部と、前記制御信号に応じて記憶デバイスに対するデータの書き込み／読み出しを行うディスク制御部と、前記各チャネル制御部からアクセス可能な共有メモリを備えるストレージ制御装置の制御方法であって、

第１の前記チャネル制御部が、前記情報処理装置と通信することにより前記情報処理装置の動作を監視し、

第１の前記チャネル制御部が、前記情報処理装置における障害を検知した場合に当該情報処理装置を特定可能な識別子を前記共有メモリに記憶し、

第２の前記チャネル制御部が、適宜なタイミングで前記共有メモリにアクセスすることにより前記識別子を取得し、

前記第２のチャネル制御部が、取得した前記識別子を有する前記情報処理装置

から送信された前記データ入出力要求を受信した場合に前記データ入出力要求に応じて実行される処理を制限するための処理を実行すること、

を特徴とするストレージ制御装置の制御方法。

【請求項３】　情報処理装置から送信されるデータ入出力要求を受信して、前記データ入出力要求に応じて記憶デバイスに対するデータの書き込み／読み出しを制御するための制御信号を出力する、互いに通信可能に接続される複数のチャネル制御部と、前記制御信号に応じて記憶デバイスに対するデータの書き込み／読み出しを行うディスク制御部と、前記各チャネル制御部からアクセス可能な共有メモリを備えるストレージ制御装置の制御方法であって、

第１の前記チャネル制御部が、前記情報処理装置と通信することにより前記情報処理装置の動作を監視し、

第１の前記チャネル制御部が、前記情報処理装置における障害を検知した場合に当該情報処理装置を特定可能な識別子を前記共有メモリに記憶するとともに前記障害が生じている旨を第２の前記チャネル制御部に通知し、

前記第２のチャネル制御部が、前記通知を受信したのに応じて前記共有メモリにアクセスすることにより前記識別子を取得し、

前記第２のチャネル制御部が、取得した前記識別子を有する前記情報処理装置から送信された前記データ入出力要求を受信した場合に前記データ入出力要求に応じて実行される処理を制限するための処理を実行すること、

を特徴とするストレージ制御装置の制御方法。

【請求項４】　請求項１に記載のストレージ制御装置の制御方法において、

前記データ入出力要求に応じて実行される処理を制限するための処理は、障害が検知された前記情報処理装置から送信された前記データ入出力要求を受信した場合に前記データ入出力要求に応じて出力される前記制御信号を出力しないように制御する処理であること、を特徴とするストレージ制御装置の制御方法。

【請求項５】　請求項１に記載のストレージ制御装置の制御方法において、

前記第１のチャネル制御部は、前記チャネル制御部と前記情報処理装置との間でハートビート信号を通信し、前記ハートビート信号が途絶したことをもって前記情報処理装置に障害が生じていることを検知すること、を特徴とするストレー

ジ制御装置の制御方法。

【請求項６】　請求項１乃至５のいずれかに記載のストレージ制御装置の制御方法において、

前記第１のチャネル制御部は、前記情報処理装置から送信される前記データ入出力要求としてファイル指定によるデータ入出力要求を受け付ける機能を備え、前記第２のチャネル制御部は、前記情報処理装置から送信される前記データ入出力要求としてブロック指定によるデータ入出力要求を受け付ける機能を備えること、を特徴とするストレージ制御装置の制御方法。

【請求項７】　情報処理装置から送信されるデータ入出力要求を受信して、前記データ入出力要求に応じて記憶デバイスに対するデータの書き込み／読み出しを制御するための制御信号を出力する、互いに通信可能に接続される複数のチャネル制御部と、前記制御信号に応じて記憶デバイスに対するデータの書き込み／読み出しを行うディスク制御部と、を備えるストレージ制御装置の制御方法であって、

前記情報処理装置から送信されてくる要求に応じて前記情報処理装置にファイル管理情報を送信し、

前記情報処理装置から送信されてくる前記ファイル管理情報に基づいて生成されたデータ入出力要求を受信してこれに応じたデータの書き込み／読み出しを前記記憶デバイスに対して実行し、

第１の前記チャネル制御部が、前記情報処理装置と通信することにより前記情報処理装置の動作を監視し、

前記第１のチャネル制御部が、前記情報処理装置における障害を検知した場合に、前記情報処理装置から送信されたデータ入出力要求が着信される第２の前記チャネル制御部により前記データ入出力要求に応じて実行される処理を制限するための処理を実行すること、

を特徴とするストレージ制御装置の制御方法。

【請求項８】　請求項１に記載のストレージ制御装置の制御方法において、

前記情報処理装置と通信することにより行われる前記情報処理装置の動作の監視は、前記情報処理装置と前記第１のチャネル制御部において動作しているクラ

スタソフトウェアの機能により実現されること、

を特徴とするストレージ制御装置の制御方法。

【請求項９】　請求項１に記載のストレージ制御装置の制御方法において、

前記第１のチャネル制御部はＬＡＮに接続するためのインタフェースを有し、

前記第２のチャネル制御部はＳＡＮに接続するためのインタフェースを有すること、を特徴とするストレージ制御装置の制御方法。

【請求項１０】　情報処理装置から送信されるデータ入出力要求を受信して、前記データ入出力要求に応じて記憶デバイスに対するデータの書き込み／読み出しを制御するための制御信号を出力する、互いに通信可能に接続される複数のチャネル制御部と、前記制御信号に応じて記憶デバイスに対するデータの書き込み／読み出しを行うディスク制御部と、

第１の前記チャネル制御部が、前記情報処理装置と通信することにより前記情報処理装置の動作を監視する手段と、

第１の前記チャネル制御部が、前記情報処理装置における障害を検知した場合に、前記情報処理装置から送信されたデータ入出力要求が着信される第２の前記チャネル制御部により前記データ入出力要求に応じて実行される処理を制限するための処理を実行する手段と、

を備えることを特徴とするストレージ制御装置。

【請求項１１】　請求項１０に記載のストレージ制御装置において、

前記第１のチャネル制御部は、前記情報処理装置から送信される前記データ入出力要求としてファイル指定によるデータ入出力要求を受け付ける機能を備え、

前記第２のチャネル制御部は、前記情報処理装置から送信される前記データ入出力要求としてブロック指定によるデータ入出力要求を受け付ける機能を備えること、を特徴とするストレージ制御装置。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】

この発明は、ストレージ制御装置の制御方法及びストレージ制御装置に関する。

【０００２】

【従来の技術】

ＩＴ関連産業の発達に伴い、コンピュータシステムが取り扱うデータ量が飛躍的に増加している。かかる膨大なデータを効率よく利用し管理するために、データセンタ等においては、ディスクアレイ装置等などの記憶装置と情報処理装置とを専用のネットワークで接続し、ディスクドライブ等のストレージに対する高速かつ大量なアクセスを実現するようにしたストレージシステムが構築されている。上記ネットワークとしては高速なデータ転送を実現するために、ファイバチャネルプロトコルに従った通信機器を用いて構築されるネットワークである、ＳＡＮ（Storage Area Network）が用いられるのが一般的である。また、最近では、複数の記憶装置と情報処理装置とを接続する上記ネットワークとしてＴＣＰ／ＩＰ（Transmission Control Protocol/Internet Protocol）プロトコルを用いたＬＡＮ（Local Area Network）を用い、情報処理装置からファイル指定によるアクセスを実現したＮＡＳ（Network Attached Storage）と呼ばれる装置も開発されている。

【０００３】

昨今では、ストレージシステムが取扱うデータに対する社会的な重要性が増しており、ストレージシステムには高い信頼性や可用性が求められるようになってきている。ここでストレージシステムの信頼性や可用性を向上させるための技術として、クラスタと呼ばれる技術が知られている。例えば、複数の情報処理装置でクラスタを構成し、情報処理装置間で障害の監視と障害検出時のフェールオーバを実現することで、ストレージシステムの信頼性や可用性を向上させることができる。

【０００４】

【特許文献１】

特開２００２－３５１７０３号公報

【０００５】

【発明が解決しようとする課題】

ところで、ストレージシステムにおいては、例えば、情報処理装置側で障害が

発生しているにもかかわらず記憶装置側でそのまま情報処理装置から送信されるデータ入出力要求についての処理を続けていると、記憶装置に記憶されているデータの破損やデータに管理上の不整合等の問題が生じる可能性がある。従って、ストレージシステムの信頼性や可用性をより確実に確保するためには情報処理装置側で発生した障害の影響ができるだけ記憶装置側に及ばないようにすることも重要である。

【０００６】

本発明は上記課題を鑑みてなされたものであり、ストレージシステムの信頼性や可用性を向上させることができる、ストレージ制御装置の制御方法及びストレージ制御装置を提供することを主たる目的とする。

【０００７】

【課題を解決するための手段】

上記目的を達成する本発明の主たる発明は、情報処理装置から送信されるデータ入出力要求を受信して、前記データ入出力要求に応じて記憶デバイスに対するデータの書き込み／読み出しを制御するための制御信号を出力する、互いに通信可能に接続される複数のチャネル制御部と、前記制御信号に応じて記憶デバイスに対するデータの書き込み／読み出しを行うディスク制御部と、を備えるストレージ制御装置の制御方法であって、第１の前記チャネル制御部が、前記情報処理装置と通信することにより前記情報処理装置の動作を監視し、第１の前記チャネル制御部が、前記情報処理装置における障害を検知した場合に、前記情報処理装置から送信されたデータ入出力要求が着信される第２の前記チャネル制御部により前記データ入出力要求に応じて実行される処理を制限するための処理を実行することとする。

【０００８】

前記制御信号は、例えば、直接的にもしくは後述する共有メモリやキャッシュメモリを介して間接的に、前記チャネル制御部から前記ディスク制御部に対して伝達される信号である。情報処理装置において障害が発生すると、そのことが第１のチャネル制御部により検知され、これに応じて第２の前記チャネル制御部により前記データ入出力要求に応じて実行される処理を制限するための処理が実行

される。これにより障害が生じている情報処理装置から送られてくる異常なデータ入出力要求に応じて第２のチャネル制御部が異常な処理を行うことにより記憶デバイスに記憶されているデータを破損したり、データに管理上の不整合が生じたりすることが防止され、ストレージシステムの信頼性や可用性を向上させることができる。

【０００９】

その他、本願が開示する課題、及びその解決方法は、発明の実施の形態の欄及び図面により明らかにされる。

【００１０】

【発明の実施の形態】

図１に本発明の一実施例として説明するストレージシステムの概略構成を示している。このストレージシステムは、例えば、データセンタ等において構築される。ストレージシステムは、例えば、インターネット上のＷｅｂサイトの運用、ＡＳＰサービスの提供、銀行や証券会社等におけるオンラインシステムやバッチ処理システムの運用等、各種の用途に利用される。

【００１１】

ストレージシステムは、ストレージ制御装置１００、ストレージ制御装置１００に通信可能に接続しストレージ制御装置１００に対してデータの書き込み／読み出しに関する要求（以下、データ入出力要求と称する）を送信する情報処理装置１乃至２（２００）、ストレージ制御装置１００により制御されデータの記憶領域を提供する１台以上のディスクドライブ３００（記憶デバイス）を含んで構成される。ストレージ制御装置１００は、情報処理装置２００から送信されてくるデータ入出力要求を受信して、前記データ入出力要求に応じてディスクドライブ３００に記憶されているデータの書き込みや読み出しに関する処理を行う。

【００１２】

ストレージ制御装置１００と情報処理装置１乃至２（２００）とは、ＬＡＮ（Local Area Network）４００を介して通信可能に接続されている。ＬＡＮ４００を介して行われる通信は、例えば、ＴＣＰ／ＩＰプロトコルに従って行われる。情報処理装置１乃至２（２００）とストレージ制御装置１００とは、ＳＡＮ（St

orage Area Network）５００を介して通信可能に接続されている。ＳＡＮ５００を介して行われる通信は、例えば、ファイバチャネル、ｉＳＣＳＩ、ＦＩＣＯＮ（登録商標）、ＥＳＣＯＮ（登録商標）、ＡＣＯＮＡＲＣ（登録商標）、ＦＩＢＡＲＣ（登録商標）等に従って行われる。

【００１３】

情報処理装置１乃至２（２００）は、いずれもＣＰＵ（Central Processing Unit）やメモリを備えるコンピュータである。情報処理装置１乃至２（２００）は、例えば、パーソナルコンピュータ、ワークステーション、メインフレームコンピュータ等である。情報処理装置１乃至２（２００）が備えるＣＰＵにより各種プログラムが実行されることにより、情報処理装置１乃至２（２００）が提供する様々な機能が実現される。情報処理装置２００では、オペレーティングシステムが動作している。オペレーティングシステム上では、各種のアプリケーションソフトウェアが動作している。後述するクラスタソフトウェアも、上記オペレーティングシステム上で動作するアプリケーションソフトウェアの一つである。

【００１４】

図２に情報処理装置２００の典型的なハードウェア構成を示している。情報処理装置２００は、ＣＰＵ２１１、メモリ２１２、記録媒体読取装置２１４、入力装置２１５、出力装置２１６、記憶装置２１７、ＬＡＮインタフェース２１８、ＳＡＮインタフェース２１９、等を備えて構成される。ＣＰＵ２１１は、メモリ２１２に格納されているプログラムを実行する。記録媒体読取装置２１４は、記録媒体２３０に記録されているプログラムやデータを読み取るための装置である。記録媒体２３０としては、例えば、フレキシブルディスクやＣＤ－ＲＯＭ、ＤＶＤ－ＲＯＭ、ＤＶＤ－Ｒ、ＤＶＤ－ＲＷ、半導体メモリ等が用いられる。記録媒体読取装置２１４は、情報処理装置２００に内蔵される形態とすることもできるし、外付されている形態とすることもできる。

【００１５】

記憶装置２１７は、例えばハードディスク装置やフレキシブルディスク装置、半導体記憶装置、等である。入力装置２１５は、ユーザやオペレータ等の人による情報処理装置２００へのデータ入力等のためのユーザインタフェースである。

入力装置２１５としては、例えば、キーボードやマウス等が用いられる。出力装置２１６は、例えば、ディスプレイやプリンタ等である。

【００１６】

ＬＡＮインタフェース２１８は、情報処理装置２００をＬＡＮ４００に接続するための通信インタフェースである。ＬＡＮインタフェース２１８としては、例えば、イーサネット（登録商標）に対応したネットワークカードが用いられる。ＳＡＮインタフェース２１９は、情報処理装置２００をＳＡＮ５００に接続するための通信インタフェースである。ＳＡＮインタフェース２１９としては、ファイバチャネルプロトコルに対応したネットワークカードが用いられる。通信ポート２１９１は、情報処理装置２００のＳＡＮ５００への接続口であり、例えば、ファイバチャネルの通信規格における「Ｎ＿Ｐｏｒｔ」や「ＮＬ＿Ｐｏｒｔ」である。通信ポート２１９１には、ＳＡＮ５００上のノードを特定するためのネットワークアドレスであるWWNが付与される。

【００１７】

図１に示すストレージ制御装置１００は、チャネル制御部１乃至２（１１０）、共有メモリ１２０、キャッシュメモリ１３０、ディスク制御部１４０、接続部１５０等を備えて構成される。ストレージ制御装置１００は、情報処理装置１乃至２（２００）との間の通信に関する機能を提供するチャネル制御部１乃至２（１１０）を備えている。チャネル制御部１（１１０）は、情報処理装置１乃至２（２００）との間でＬＡＮ４００を介した通信に関する機能を提供する。チャネル制御部２（１１０）は、情報処理装置１乃至２（２００）との間でＳＡＮ５００を介した通信に関する機能を提供する。チャネル制御部１（１１０）とチャネル制御部２（１１０）とは、互いにバスライン１０５によって通信可能に接続されている。チャネル制御部１（１１０）とチャネル制御部２（１１０）とは、バスライン１０５を介して互いにデータ信号や割込信号等を送信もしくは受信することができる。以下ではチャネル制御部１（１１０）をＣＨＮとも称する。また、チャネル制御部２（１１０）をＣＨＦとも称する。

【００１８】

図３にチャネル制御部１（ＣＨＮ）１１０のハードウェア構成を示している。

チャネル制御部１（ＣＨＮ）１１０は、ネットワークインタフェース部１１１、ＣＰＵ１１２、メモリ１１３、入出力制御部１１４、通信ポート１１７、等を備えている。ネットワークインタフェース部１１１は、ＬＡＮ４００に接続するための通信インタフェースである。通信ポート１１７には、ＬＡＮ４００に接続するための通信線が接続される。

【００１９】

チャネル制御部１（ＣＨＮ）１１０では、オペレーティングシステムが動作している。また、オペレーティングシステム上では、各種のアプリケーションソフトウェアが動作している。オペレーティングシステムやアプリケーションソフトウェアの機能は、ＣＰＵ１１２がメモリ１１３にロードされるプログラムが実行されることにより実現される。前記プログラムは、例えば、ディスクドライブ３００やＮＶＲＡＭ１１５等に記憶されている。

【００２０】

オペレーティングシステムは、例えば、ＵＮＩＸ（登録商標）系やＷｉｎｄｏｗｓ（登録商標）系のオペレーティングシステムでありファイルシステムを備えている。また、オペレーティングシステム上では、ＮＦＳ（Network File System）、ＣＩＦＳ（Common Internet File System）等のファイルシステムプロトコルが動作しており、情報処理装置１乃至２からファイル指定によるデータ入出力要求（以下、ファイルアクセス要求と称する）を受け付ける。すなわち、チャネル制御部１（ＣＨＮ）１１０は、情報処理装置１乃至２等のＬＡＮ４００に接続する装置に対してＮＡＳ（Network Attached Storage）として機能している。

【００２１】

入出力制御部１１４は、ディスク制御部１４０やキャッシュメモリ１３０、共有メモリ１２０、との間でデータやコマンドの授受を行う。入出力制御部１１４は、Ｉ／Ｏプロセッサ（Input/Output）１１９及びＮＶＲＡＭ（Non Volatile RAM）１１５を備えている。Ｉ／Ｏプロセッサ１１９は、例えば、１チップのマイコンで構成される。Ｉ／Ｏプロセッサ１１９は、例えば、ＤＭＡ（Direct Memory Access）プロセッサである。Ｉ／Ｏプロセッサ１１９は、上記データやコマンドの授受を制御し、ＣＰＵ１１２と接続部（スイッチ）との間の通信を中継して

いる。ＮＶＲＡＭ１１５は、Ｉ／Ｏプロセッサ１１９の制御を司るプログラムが格納される不揮発性メモリである。

【００２２】

図４にチャネル制御部２（ＣＨＦ）１１０のハードウェア構成を示している。チャネル制御部２（ＣＨＦ）１１０は、ネットワークインタフェース部１１１、メモリ１１３、入出力制御部１１４、Ｉ／Ｏプロセッサ１１９、ＮＶＲＡＭ１１５、通信ポート１１７を備える。

【００２３】

チャネル制御部２（ＣＨＦ）１１０は、ＣＰＵ１１２を備えていない点でチャネル制御部１（ＣＨＮ）１１０と構成が異なる。チャネル制御部２（ＣＨＦ）１１０ではファイルシステムは動作しておらず、情報処理装置１乃至２（２００）からファイル指定ではなくブロック指定によるデータ入出力要求（以下、ブロックアクセス要求と称する）を受け付ける。また、チャネル制御部１（ＣＨＮ）１１０がＬＡＮ４００を介して行われる通信に関する機能を提供するのに対し、チャネル制御部２（ＣＨＦ）１１０はＳＡＮ５００を介して行われる通信に関する機能を提供する。

【００２４】

ネットワークインタフェース部１１１は、ＳＡＮ５００に接続するための通信インタフェースである。ネットワークインタフェース部１１１は、２つの通信ポート１１７を備えている。各通信ポート１１７には、ＳＡＮ５００に接続するための通信線が接続される。各通信ポート１１７には、それぞれＳＡＮ５００におけるネットワークアドレスが付与される。例えば、ＳＡＮ５００の通信プロトコルがファイバチャネルである場合には、前記ネットワークアドレスは、WWN（World Wide Name）であり、各通信ポート１１７にはそれぞれ異なるWWNが付与される。

【００２５】

入出力制御部１１４は、ディスク制御部１４０、キャッシュメモリ１３０、共有メモリ１２０、とそれぞれとの間でデータやコマンドの授受を行う。入出力制御部１１４は、Ｉ／Ｏプロセッサ（Input/Output）１１９及びＮＶＲＡＭ（Non

Volatile RAM）１１５を備えている。Ｉ／Ｏプロセッサ１１９は、例えば、１チップのマイコンで構成される。Ｉ／Ｏプロセッサ１１９は、例えば、ＤＭＡ（Direct Memory Access）プロセッサである。Ｉ／Ｏプロセッサ１１９は、上記データやコマンドの授受を制御し、ＣＰＵ１１２と接続部（スイッチ）１５０との間の通信を中継する。ＮＶＲＡＭ１１５は、Ｉ／Ｏプロセッサ１１９の制御を司るプログラムが格納される不揮発性メモリである。

【００２６】

ディスク制御部１４０は、ディスクドライブ３００の制御を行う。ディスク制御部１４０は、例えば、チャネル制御部１１０が情報処理装置２００から受信したデータ書き込みコマンドに従ってディスクドライブ３００へデータの書き込みを行う。図５にディスク制御部１４０のハードウェア構成を示している。ディスク制御部１４０は、インタフェース部１４１、メモリ１４３、ＣＰＵ１４２、ＮＶＲＡＭ１４４、等を備えている。インタフェース部１４１は、接続部１５０を介してチャネル制御部１乃至２（１１０）との間での通信や、共有メモリ１２０、キャッシュメモリ１３０へのアクセスを行うための通信インタフェースや、ディスクドライブ３００との間で通信を行うための通信インタフェースを備えている。

【００２７】

ＣＰＵ１４２は、ディスク制御部１４０全体の制御を司ると共に、チャネル制御部１乃至２（１１０）やディスクドライブ３００との間の通信を行う。メモリ１４３やＮＶＲＡＭ１４４に格納された各種プログラムを実行することにより本実施の形態に係るディスク制御部１４０の機能が実現される。ディスク制御部１４０により実現される機能としては、例えば、ディスクドライブ３００の制御やＲＡＩＤの機能に関する各種制御等がある。ＮＶＲＡＭ１４４は、ＣＰＵ１４２の制御を司るプログラムを格納する不揮発性メモリである。

【００２８】

図１に示すように、チャネル制御部１乃至２（１１０）、共有メモリ１２０、キャッシュメモリ１３０、ディスク制御部１４０は、それぞれ接続部１５０を介して互いに通信可能に接続されている。接続部１５０は、例えば、高速スイッチ

ングによりデータ伝送を行う超高速クロスバスイッチなどのスイッチである。接続部１５０は、スイッチングに関する制御を行うための制御プロセッサ（不図示）を備えている。共有メモリ１２０及びキャッシュメモリ１３０は、チャネル制御部１乃至２（１１０）、ディスク制御部１４０により共有されるメモリである。共有メモリ１２０は、主に制御情報やコマンド等を記憶するために利用されるのに対し、キャッシュメモリ１３０は、主にデータを記憶するために利用される。

【００２９】

＝＝＝基本動作＝＝＝

次に、情報処理装置１乃至２（２００）から送信されてくるデータ書き込み要求やデータ読み出し要求などのデータ入出力要求を受信した場合におけるストレージ制御装置１００の基本動作について説明する。

【００３０】

まず、情報処理装置１乃至２（２００）からストレージ制御装置１００に対してデータ書き込み要求が送信された場合について説明する。ストレージ制御装置１００のチャネル制御部１乃至２（１１０）は、情報処理装置１乃至２（２００）から送られてくるデータ書き込み要求を受信すると、データ書き込みコマンドを共有メモリ１２０に書き込むと共に、情報処理装置１乃至２（２００）から受信した書き込みデータをキャッシュメモリ１３０に書き込む。チャネル制御部１乃至２（１１０）は、キャッシュメモリ１３０に対するデータの書き込みが完了すると、情報処理装置２００に書き込み完了報告を送信する。このように情報処理装置２００への完了報告は、ディスクドライブ３００への実際のデータの書き込み動作とは非同期に行われる。ディスク制御部１４０はリアルタイム（例えば、一定の時間間隔で）に共有メモリ１２０の内容を監視している。ディスク制御部１４０は、上記監視により共有メモリ１２０にデータ書き込みコマンドが書き込まれていることを検知すると、キャッシュメモリ１３０から書き込み対象となるデータ（以下、書き込みデータと称する）を読み出して、読み出した書き込みデータをディスクドライブ３００に書き込む。以上のようにしてデータ書き込み要求に対応したディスクドライブ３００へのデータの書き込みが行われる。

【００３１】

次に、情報処理装置２００からストレージ制御装置１００に対してデータ書き込み要求が送信された場合におけるストレージ制御装置１００の基本的な動作について説明する。ストレージ制御装置１００は、情報処理装置２００から送られてくるデータ読み出し要求を受信すると、この要求に対応するデータ読み出しコマンドをディスク制御部１４０に送出する。なお、チャネル制御部１乃至２（１１０）からディスク制御部１４０へのデータ読み出しコマンドの伝達は、共有メモリ１２０を介して行われることもある。

【００３２】

ディスク制御部１４０は、チャネル制御部１乃至２（１１０）からデータ読み出しコマンドを受領すると、そのコマンドに指定されている読み出し対象のデータをディスクドライブ３００から読み出して、読み出したデータをキャッシュメモリ１３０に書き込む。ディスク制御部１４０は、キャッシュメモリ１３０へのデータ転送が完了すると、その旨をチャネル制御部１乃至２（１１０）に通知する。そして前記通知を受信したチャネル制御部１乃至２（１１０）は、キャッシュメモリ１３０に記憶されている読み出し対象のデータを情報処理装置１乃至２（２００）に転送する。

【００３３】

上述したように、チャネル制御部１（ＣＨＮ）１１０は、情報処理装置１乃至２（２００）からデータ入出力要求としてＬＡＮ４００を介してファイル指定によるファイルアクセス要求を受け付ける。一方、チャネル制御部２（ＣＨＦ）１１０は、情報処理装置１乃至２（２００）からデータ入出力要求としてＳＡＮ５００を介してブロック指定によるブロックアクセス要求を受け付ける。

【００３４】

＝＝＝障害監視と障害対応＝＝＝

チャネル制御部１（ＣＨＮ）１１０と情報処理装置１乃至２（２００）とにおいては、これら装置間で互いに動作状態を監視し合う機能を提供するアプリケーションソフトウェアが動作している。なお、近年、複数のコンピュータを含んで構成されるシステムにおいては、可用性（ＨＡ（High Availability））の向上

、負荷分散（ロードバランシング）による処理効率の向上、同一処理の並列実行による信頼度を向上等を目的として、いわゆるクラスタの仕組みが導入されることが多い。このような場合には、お互いに他の装置の動作状態を監視する上記の機能は、クラスタの機能を実現するためのアプリケーションソフトウェア（以下、クラスタソフトウェア１６０と称する）の機能として提供される。

【００３５】

クラスタソフトウェア１６０は、ＬＡＮ４００を介してお互いにハートビートメッセージを送受信することで、相手方装置が正常に動作しているかどうかを監視している。クラスタソフトウェア１６０は、期待される時刻に相手方の装置から送られてくるハートビートメッセージを受信できている場合には、当該相手方装置は正常に動作しているものと判断する。また、クラスタソフトウェア１６０は、期待される時刻に相手の装置から送られてくるハートビートメッセージを受信できていない場合、すなわち、ハートビートメッセージが途絶している場合には、相手装置に何らかの障害が発生しているものと判断する。なお、上記にいう期待される時刻は、ハートビートメッセージがストレージ制御装置１００、情報処理装置２００、ＬＡＮ４００等に与える負荷や障害検知に要求される迅速性等を考慮して、適切な値に設定される。チャネル制御部１（２００）において動作しているクラスタソフトウェア１６０は、ハートビートメッセージを送信してくる相手装置側のＩＰアドレスを記憶している。クラスタソフトウェア１６０は、ハートビートメッセージが送信されてこない情報処理装置２００が存在する場合に、その情報処理装置２００のＩＰアドレスを特定する機能を備えている。なお、他の装置の動作状態を監視する仕組みは、ハートビートメッセージを用いる上述のものに限られない。

【００３６】

ストレージ制御装置１００は、以上に説明したクラスタソフトウェア１６０により提供される障害監視の仕組みによって、ある情報処理装置２００からのハートメッセージが途絶していることを検知した場合には、その情報処理装置２００（以下、障害中の情報処理装置２００と称する）からチャネル制御部２（１１０）に対して行われるデータ入出力要求に応じて実行される処理を制限するための

処理を実行する。ここでデータ入出力要求に応じて実行される処理を制限するための処理とは、例えば、通常の運用状態においてそのデータ入出力要求に対応して実行されるようにされている処理を実行しないようにすることをいう。具体的には、例えば、チャネル制御部２（１１０）が障害中の情報処理装置２００から送信されたデータ入出力要求を受信した場合に、そのデータ入出力要求に対応する処理を実行しないようにする処理、前記データ入出力要求に応じて出力される前記制御信号を出力しないように制御する処理、またはそのデータ入出力要求に対する情報処理装置２００への受信通知を送信しないようにする処理をいう。以下では、チャネル制御部１（ＣＨＮ）１１０が、ある情報処理装置２００からのハートビートメッセージが途絶していることを検知した場合に、前記情報処理装置２００からチャネル制御部２（ＣＨＦ）１１０に対して送信されたデータ入出力要求に応じて実行される処理を制限する処理について説明する。

【００３７】

図６は、ある情報処理装置２００からのハートビートメッセージが途絶していることを検知した場合に、その情報処理装置２００からチャネル制御部２（ＣＨＦ）１１０に対して行われるデータ入出力要求に応じて実行される処理が制限される処理の一例を説明するフローチャートである。チャネル制御部１（ＣＨＮ）１１０で動作するクラスタソフトウェア１６０は、ある情報処理装置２００において障害が発生し（S610）、その情報処理装置２００からのハートビートメッセージが途絶していることを検知すると（S611）、障害中の情報処理装置２００に付与されているＳＡＮ５００上のネットワークアドレス（WWN）を検索する。ここでこの検索は、障害中の情報処理装置２００に付与されているＩＰアドレスを特定し、そのＩＰアドレスを検索キーとして、図７に示す情報処理装置－WWN対応管理テーブル７００を参照することにより行われる。情報処理装置－WWN対応管理テーブル７００は、チャネル制御部１（ＣＨＮ）１１０のメモリ１１３、ＮＶＲＡＭ１１５、もしくは、ディスクドライブ３００に記憶されているテーブルである。情報処理装置－WWN対応管理テーブル７００には、各情報処理装置２００に付与されているＬＡＮ４００上のＩＰアドレスと、情報処理装置２００に付与されているＳＡＮ５００上のWWNとの対応づけが登録されている。

クラスタソフトウェア１６０は、検索したWWNをメモリ１１３に書き込む（S612）。

【００３８】

次にクラスタソフトウェア１６０は、メモリ１１３上に管理されている書き込み指示フラグをＯＮに設定する（S613）。チャネル制御部１（ＣＨＮ）１１０のＩ／Ｏプロセッサ１１９は、書き込み指示フラグの内容をリアルタイムに監視している。チャネル制御部１（ＣＨＮ）１１０のＩ／Ｏプロセッサ１１９は、書き込み指示フラグがＯＮになっていることを検知すると（S614:YES）、メモリ１１３に書き込まれているWWNを共有メモリ１２０に転送する（S615）。ここで共有メモリ１２０に転送された前記WWNは、共有メモリ１２０に確保されている記憶領域であるWWN書き込みエリア１２０１に書き込まれる。

【００３９】

一方、チャネル制御部２（ＣＨＦ）１１０のＩ／Ｏプロセッサ１１９は、共有メモリ１２０のWWN書き込みエリア１２０１の内容をリアルタイムに監視している（S616）。Ｉ／Ｏプロセッサ１１９は、前記監視により新たなWWNがWWN書き込みエリア１２０１に書き込まれていることを検知すると（S617:YES）、そのWWNを読み出して、チャネル制御部２（ＣＨＦ）１１０のメモリ１１３に転送する（S618）。これによりチャネル制御部２（ＣＨＦ）１１０のメモリ１１３には、障害中の情報処理装置２００のWWNがリアルタイムに管理される。

【００４０】

チャネル制御部２（ＣＨＦ）１１０のＩ／Ｏプロセッサ１１９は、チャネル制御部２（ＣＨＦ）１１０のメモリ１１３に前記WWNが記憶されている場合には、情報処理装置１乃至２（２００）からＳＡＮ５００を介して送信されてくるデータ入出力要求についてそのデータ入出力要求に応じて実行される処理を制限するかどうかの判定を行う。そして、チャネル制御部２（ＣＨＦ）１１０は、メモリ１１３に記憶されているWWNと同じWWNが付与されている情報処理装置２００からデータ入出力要求が送信されてくると、そのデータ入出力要求についての処理を制限する（S619）。

【００４１】

ここでデータ入出力要求がメモリ１１３に記憶されているWWNが付与された情報処理装置２００から送信されたものであるかどうかの判断は、メモリ１１３に記憶されているWWNと、送信元の通信ポート２１９１を特定するためにデータ入出力要求に付帯して送られてくるWWNとを対照することにより行われる。

【００４２】

以上の仕組みによれば、情報処理装置２００において障害が発生した場合には、その情報処理装置２００から送信されたデータ入出力要求に関する処理が制限される。これにより、情報処理装置２００から送られてくる異常なコマンドや異常なデータによりディスクドライブ３００に記憶されるデータが破損したり、データに管理上の不整合が生じたりすることを防ぐことができ、ストレージシステムの信頼性や可用性を向上させることができる。

【００４３】

なお、情報処理装置２００の障害が復旧した場合には、クラスタソフトウェア１６０の機能により、もしくは、オペレータ等による手動操作により、該当の情報処理装置２００のWWNが、WWN書き込みエリア１２０１から削除される。そして、Ｉ／Ｏプロセッサ１１９は、前記監視において、チャネル制御部２（１１０）のメモリ１１３に記憶されているWWNが、共有メモリ１２０のWWN書き込みエリア１２０１から削除されている場合には、チャネル制御部２（１１０）のメモリ１１３からそのWWNを削除する。これにより情報処理装置２００の障害が復旧して正常に動作を開始した場合には、制限されていた該当の情報処理装置２００から送信されたデータ入出力要求についての処理が自動的に再開される。

【００４４】

なお、例えば、WWN書き込みエリア１２０１に記憶されている各WWNに対応させたフラグ（以下、抑止解除フラグと称する）を共有メモリ１２０に管理し、チャネル制御部２（ＣＨＦ）１１０が抑止解除フラグの内容に応じてデータ入出力要求に応じた処理を制限する処理を実行するかどうかを決定するようにしてもよい。このように抑止解除フラグを用いる場合には、例えば、各チャネル制御部２（ＣＨＦ）が参照したかどうかを示す情報を共有メモリ１２０に管理してお

き、ＷＷＮを参照すべき全てのチャネル制御部２（ＣＨＦ）１１０が前記ＷＷＮを参照したことを確認した後に前記ＷＷＮをＷＷＮ書き込みエリア１２０１から削除するようにしてもよい。このようにすることで、チャネル制御部２（ＣＨＦ）１１０が複数存在する場合において各チャネル制御部２（ＣＨＦ）１１０が個別に上記データ入出力要求に応じた処理を制限する必要があるかどうかを判断できるようにすることができる。

【００４５】

図８はある情報処理装置２００からのハートビートメッセージが途絶していることを検知した場合に、その情報処理装置２００からチャネル制御部２（１１０）に対して送信されてくるデータ入出力要求に応じて実行される処理が制限される処理の他の実施の形態を説明するフローチャートである。チャネル制御部１（ＣＨＮ）１１０で動作するクラスタソフトウェア１６０は、ある情報処理装置２００において障害が発生し（S810）、その情報処理装置２００からのハートビートメッセージが途絶していることを検知すると（S811）、いずれかの情報処理装置２００に障害が発生していることを示す割込信号をバスライン１０５を介してチャネル制御部２（１１０）に出力する（S812）。また、クラスタソフトウェア１６０は、障害中の情報処理装置２００に付与されているＳＡＮ５００上のネットワークアドレス（ＷＷＮ）を検索する。ここでこの検索は、障害中の情報処理装置２００に付与されているＩＰアドレスを特定し、そのＩＰアドレスを検索キーとして、情報処理装置－ＷＷＮ対応管理テーブル７００を参照することにより行われる。クラスタソフトウェア１６０は、検索したＷＷＮをメモリ１１３に書き込む（S813）。

【００４６】

次にクラスタソフトウェア１６０は、メモリ１１３上に管理されている書き込み指示フラグをＯＮに設定する（S814）。チャネル制御部１（ＣＨＮ）１１０のＩ／Ｏプロセッサ１１９は、書き込み指示フラグの内容をリアルタイムに監視している。チャネル制御部１（ＣＨＮ）１１０のＩ／Ｏプロセッサ１１９は、書き込み指示フラグがＯＮになっていることを検知すると（S815:YES）、メモリ１１３に書き込まれているＷＷＮを共有メモリ１２０に転送する（S815）。ここで共

有メモリ１２０に転送された前記WWNは、共有メモリ１２０に確保されている記憶領域であるWWN書き込みエリア１２０１に書き込まれる。

【００４７】

一方、チャネル制御部２（ＣＨＦ）１１０のＩ／Ｏプロセッサ１１９は、チャネル制御部１（ＣＨＮ）１１０から出力された割込信号を受信したのに応じて（S817）共有メモリ１２０にアクセスし（S818）、WWN書き込みエリア１２０１に書き込まれている新たなWWNをチャネル制御部２（ＣＨＦ）１１０のメモリ１１３に記憶する（S819）。これによりチャネル制御部２（ＣＨＦ）１１０のメモリ１１３には、障害中の情報処理装置２００のWWNがリアルタイムに管理される。

【００４８】

チャネル制御部２（ＣＨＦ）１１０のＩ／Ｏプロセッサ１１９は、チャネル制御部２（ＣＨＦ）１１０のメモリ１１３に前記WWNが記憶されている場合には、情報処理装置１乃至２（２００）からＳＡＮ５００を介して送信されてくるデータ入出力要求についてそのデータ入出力要求に応じて実行される処理を制限するかどうかの判定を行う。そして、チャネル制御部２（ＣＨＦ）１１０は、メモリ１１３に記憶されているWWNと同じWWNが付与されている情報処理装置２００からデータ入出力要求が送信されてきた場合にはそのデータ入出力要求に応じて実行される処理を制限する（S820）。ここでデータ入出力要求がメモリ１１３に記憶されているWWNが付与された情報処理装置２００から送信されたものであるかどうかの判断は、メモリ１１３に記憶されているWWNと、送信元の通信ポート２１９１を特定するためにデータ入出力要求に付帯して送られてくるWWNとを対照することにより行われる。

【００４９】

以上の仕組みによれば、チャネル制御部１（ＣＨＮ）１１０が、情報処理装置２００における障害を検知した場合に、前記情報処理装置２００から送信されたデータ入出力要求が着信されるチャネル制御部２（ＣＨＦ）１１０により前記データ入出力要求に応じて実行される処理が制限されるように制御される。これにより障害が生じている情報処理装置２００から送られてくる異常なデータ入出力

要求に応じてチャネル制御部２（ＣＨＦ）１１０が異常な処理を行うことによりディスクドライブ３００に記憶される（もしくは、記憶されている）データが破損したり、データに管理上の不整合が生じたりすることが防止され、ストレージシステムの信頼性や可用性を向上させることができる。また、チャネル制御部１（１１０）からチャネル制御部２（１１０）に対して情報処理装置２００で障害が発生した場合にバスライン１０５を介して割込信号が出力され、チャネル制御部２（ＣＨＦ）１１０は前記割込信号が入力されたのに応じて共有メモリ１２０にアクセスするので、リアルタイムに共有メモリ１２０にアクセスする上述の図６に示す仕組みに比べてチャネル制御部２（ＣＨＦ）１１０の負荷も少なくて済む。

【００５０】

また、このようにストレージ制御装置１００がその内部に複数のチャネル制御部１１０を備え、そのうちの少なくとも一つのチャネル制御部１１０が、情報処理装置２００における障害を検知する仕組みを提供することができる構成である場合には、ストレージ制御装置１００内部の通信により障害中の情報処理装置２００から送信されたデータ入出力要求に応じて実行される処理を制限する仕組みを、ストレージ制御装置１００の内部に持たせることが可能であり、信頼性及び可用性に優れたストレージ制御装置を提供することが可能となる。

【００５１】

＝＝＝高速アクセス制御＝＝＝

次に本発明の他の実施の形態に係るファイルの高速アクセス制御について説明する。本実施の形態に係るファイルの高速アクセス制御は、情報処理装置２００から、ディスクドライブ３００に記憶されているファイルデータに対して、ＳＡＮ５００を介したブロック単位の高速なデータアクセスを行うための制御である。　図１に示すように、情報処理装置１（２００）は、ＬＡＮ４００を介してチャネル制御部１（ＣＨＮ）１１０と接続されている。また情報処理装置１（２００）は、ＳＡＮ５００を介してチャネル制御部２（ＣＨＦ）１１０とも接続されている。これにより情報処理装置１（２００）は、チャネル制御部１（ＣＨＮ）１１０を通じてディスクドライブ３００に記憶されているファイルデータをアク

セスすることもできるし、チャネル制御部２（ＣＨＦ）１１０を通じて上記同一データをアクセスすることもできる。但し、チャネル制御部１（ＣＨＮ）１１０を介してアクセスする場合はファイル単位でのアクセスとなるのに対し、チャネル制御部２（ＣＨＦ）１１０を介してアクセスする場合はブロック単位でのアクセスとなる。

【００５２】

通常、情報処理装置１（２００）が、チャネル制御部１（ＣＨＮ）１１０を介してディスクドライブ３００に記憶されているデータにアクセスする場合には、チャネル制御部１（ＣＨＮ）１１０に対してファイル名を指定したファイルアクセス要求を行うが、本実施の形態に係るファイルの高速アクセス制御によりディスクドライブ３００に記憶されているデータにアクセスする場合には、情報処理装置１（２００）は、まずチャネル制御部１（ＣＨＮ）１１０に対してファイル名を指定して、ファイルについての記憶装置の記憶領域上の記憶位置を特定する情報であるメタデータ（ファイル管理情報）の要求（リクエスト）を行う。メタデータは、例えば、ＵＮＩＸ（登録商標）における「i-node」である。メタデータの要求を受け付けたチャネル制御部１（ＣＨＮ）１１０は、メモリ１１３またはキャッシュメモリ１３０に記憶されている当該ファイル名に対応するメタデータを読み出す。そして、読み出したメタデータをＬＡＮ４００を介して情報処理装置１（２００）に送信する。なお、メタデータはディスクドライブ３００にも記憶されているので、チャネル制御部１（ＣＨＮ）１１０はディスクドライブ３００からメタデータを読み出すようにすることもできる。

【００５３】

情報処理装置（２００）はメタデータを取得することにより、当該ファイルの記憶位置やデータサイズを知ることができる。情報処理装置（２００）はこれらの情報に基づいてファイルデータに対するブロックアクセス要求を生成する。そして当該ブロックアクセス要求をＳＡＮ５００を介してチャネル制御部２（ＣＨＦ）１１０に対して送信する。

【００５４】

チャネル制御部２（ＣＨＦ）１１０は、ネットワークインタフェース部１１１

により上記ブロックアクセス要求を受け付ける。そしてＩ／Ｏプロセッサ１１９は当該データの記憶位置とデータ長等を抽出し、上記ブロックアクセス要求に対応するＩ／Ｏ要求を生成してディスク制御部１４０に出力する。このようにしてデータの読み出しや書き込み等が行われる。

【００５５】

ＳＡＮ５００はＬＡＮ４００と比較して高速なデータ転送が可能なネットワークであるので、ディスクドライブ３００に記憶されているファイルデータに高速にアクセスすることができる。

情報処理装置１（２００）は、ディスクドライブ３００からファイルデータを読み出す場合には、チャネル制御部２（ＣＨＦ）１１０に対して当該ファイルデータのアドレスとサイズを指定してブロック単位でのデータ読み出し要求を送信する。チャネル制御部２（ＣＨＦ）１１０はディスクドライブ３００から読み出したデータをＳＡＮ５００を介して情報処理装置１（２００）に送信する。情報処理装置１（２００）は、データをチャネル制御部２（ＣＨＦ）１１０から取得したら読み出し処理を終了する。なお、チャネル制御部１（ＣＨＮ）１１０からメタデータを取得する際に当該ファイルに対してロックを掛けていた場合には、チャネル制御部１（ＣＨＮ）１１０に対してロック解除要求を送信する。

【００５６】

一方、ディスクドライブ３００にファイルデータを書き込む場合には、情報処理装置１（２００）は、当該書き込みデータと共に書き込みデータのアドレスとサイズを指定してブロック単位のデータ書き込み要求をチャネル制御部２（ＣＨＦ）１１０に対して送信する。チャネル制御部２（ＣＨＦ）１１０は当該書き込みデータをディスクドライブ３００に書き込み、書き込み完了メッセージを情報処理装置１（２００）に送信する。情報処理装置１（２００）はチャネル制御部２（ＣＨＦ）１１０から書き込み完了のメッセージを受信したら、チャネル制御部１（ＣＨＮ）１１０に対してメタデータの更新を要求する。

【００５７】

本実施の形態に係るファイルの高速アクセス制御は、データサイズの大きなファイルにアクセスする場合に効果が大きい。データサイズの大きなファイルへの

アクセスを高速なＳＡＮ５００を介して行うことにより、ファイルデータに対する読み出し、書き込みの時間を短縮できる。これは本実施の形態に係るストレージシステム６００においては、ストレージ制御装置１００のスロット内にチャネル制御部１（ＣＨＮ）１１０、チャネル制御部２（ＣＨＦ）１１０を混在させて装着することが可能であり、チャネル制御部１（ＣＨＮ）１１０を介したデータアクセスとチャネル制御部２（ＣＨＦ）１１０を介したデータアクセスのそれぞれの特長をうまく利用することができる場合に実現できるのである。

【００５８】

本実施の形態にかかるファイルの高速アクセス制御は、情報処理装置１（２００）からチャネル制御部１（ＣＨＮ）１１０に対するアクセスと、情報処理装置１（２００）からチャネル制御部２（ＣＨＦ）１１０に対するアクセスとが、関連をもって行われる。上述した障害監視と障害対応の仕組みは、このような形態においても実現される。つまり、情報処理装置１（２００）とクラスタを構成しているチャネル制御部１（ＣＨＮ）１１０が、情報処理装置１（２００）において生じている障害を検知すると、チャネル制御部２（ＣＨＦ）１１０により前記データ入出力要求に応じて実行される処理が制限されるように制御される。すなわち、このように情報処理装置１（２００）からチャネル制御部１（ＣＨＮ）１１０に対するアクセスと、情報処理装置１（２００）からチャネル制御部２（ＣＨＦ）１１０に対するアクセスとが、関連をもって行われる場合に、上述した障害監視と障害対応の仕組みは、特に有用である。これにより、障害が生じている情報処理装置１（２００）から送られてくる異常なデータ入出力要求に応じてチャネル制御部２（ＣＨＦ）１１０が異常な処理を行うことによりディスクドライブ３００に記憶される（もしくは、記憶されている）データが破損したり、データに管理上の不整合が生じたりすることが防止され、高速アクセス制御が行われた場合におけるストレージシステムの信頼性や可用性を向上させることができる。

【００５９】

＝＝＝ストレージシステムの他の形態＝＝＝

図９は、ストレージシステムの他の形態を示している。このストレージシステ

ムは、ストレージ制御装置１００、ストレージ制御装置１００に通信可能に接続しストレージ制御装置１００に対してデータ入出力要求を送信する情報処理装置１乃至５（２００）、ストレージ制御装置１００により制御されデータの記憶領域を提供する記憶デバイスである一台以上のディスクドライブ３００等を含んで構成される。ストレージ制御装置１００は、情報処理装置１乃至５（２００）から送信されたデータ入出力要求を受信して、このデータ入出力要求に応じてディスクドライブ３００に記憶されているデータの書き込みや読み出しに関する処理を行う。

【００６０】

ストレージ制御装置１００と情報処理装置１乃至５（２００）とは、ＬＡＮ（Local Area Network）４００を介して通信可能に接続されている。ＬＡＮ４００を介して行われる通信は、例えば、ＴＣＰ／ＩＰプロトコルに従って行われる。また、情報処理装置３乃至４（２００）とストレージ制御装置１００とは、ＳＡＮ（Storage Area Network）５００を介して通信可能に接続されている。ＳＡＮ５００を介して行われる通信はファイバチャネルに従って行われる。情報処理装置５（２００）とストレージ制御装置１００とは、ｉＳＣＳＩ、ＦＩＣＯＮ（登録商標）やＥＳＣＯＮ（登録商標）、ＡＣＯＮＡＲＣ（登録商標）、ＦＩＢＡＲＣ（登録商標）等の通信プロトコルに従って通信が行われる通信経路５５０を介して通信可能に接続されている。

【００６１】

情報処理装置１乃至５（２００）は、それぞれＣＰＵ（Central Processing Unit）やメモリを備えるコンピュータである。情報処理装置１乃至５（２００）は、例えば、パーソナルコンピュータ、ワークステーション、メインフレームコンピュータ等である。情報処理装置１乃至５（２００）が備えるＣＰＵにより各種プログラムが実行されることにより、情報処理装置１乃至５（２００）が提供する様々な機能が実現される。また、情報処理装置１乃至５（２００）では、オペレーティングシステムが動作している。また、オペレーティングシステム上では、各種のアプリケーションプログラムが動作しており、情報処理装置１乃至５（２００）のうちの少なくともいずれかにおいては上述のクラスタソフトウェア

１６０が動作している。

【００６２】

ストレージ制御装置１００は、情報処理装置１乃至５（２００）との間の通信に関する機能を提供する、チャネル制御部１乃至８（１１０）を備えている。このうちチャネル制御部１乃至４（ＣＨＮ）１１０は、情報処理装置１乃至３（２００）との間でＬＡＮ４００を介した通信に関する機能を提供する。なお、チャネル制御部１乃至８（１１０）は、バスライン１０５によって互いに通信可能に接続されている。チャネル制御部１乃至４（１１０）は、このバスライン１０５を介して互いにデータ信号や割込信号等を送信もしくは受信することができる。

【００６３】

チャネル制御部１乃至４（１１０）は、第一の実施例で説明したチャネル制御部１（ＣＨＮ）１１０と同等のハードウェア及びソフトウェア構成を備える。また、チャネル制御部５乃至８（ＣＨＦ）１１０は、第一の実施例で説明したチャネル制御部２（ＣＨＦ）１１０と同等のハードウェア及びソフトウェア構成を備える。

【００６４】

チャネル制御部１乃至４（ＣＨＦ）１１０のうちの少なくともいずれかにおいては、上述のクラスタソフトウェア１６０が動作している。クラスタソフトウェア１６０は、ある情報処理装置２００からのハートビートメッセージが途絶していることを検知した場合に、その情報処理装置２００から他のチャネル制御部１乃至８（１１０）に対して行われるデータ入出力要求に応じて実行される処理を制限させる、図６もしくは図８のフローチャートで説明した処理に相当する処理を実行する。なお、図９に示すストレージシステムの場合には、チャネル制御部１（ＣＨＮ）１１０がある情報処理装置２００における障害を検知した場合にチャネル制御部２（ＣＨＦ）１１０に対してのデータ入出力要求に応じて実行される処理が制限されるだけでなく、当該チャネル制御部１（ＣＨＮ）１１０以外の他のチャネル制御部１（ＣＨＮ）１１０に対して送信されるデータ入出力要求に応じて実行される処理についても制限されるようにすることもできる。

【００６５】

ディスク制御部１乃至４（１４０）のハードウェア及びソフトウェア構成は、第一の実施例で説明したディスク制御部１４０と同等の構成である。チャネル制御部１乃至８（１１０）、共有メモリ１２０、キャッシュメモリ１３０、ディスク制御部１４０は、接続部１５０を介して通信可能に接続されている。接続部１５０は、例えば、高速スイッチングによりデータ伝送を行う超高速クロスバスイッチなどのスイッチである。共有メモリ１２０及びキャッシュメモリ１３０は、チャネル制御部１１０、ディスク制御部１４０により共有されるメモリである。共有メモリ１２０は、主に制御情報やコマンド等を記憶するために利用されるのに対し、キャッシュメモリ１３０は、主にデータを記憶するために利用される。

【００６６】

ディスクドライブ３００は、情報処理装置２００に対して記憶領域を提供する。データは、ディスクドライブ３００により提供される物理的な記憶領域上に論理的に設定される記憶領域である論理ボリュームに記憶されている。ディスクドライブ３００としては、例えば、ハードディスク装置や半導体記憶装置等、様々なものを用いることができる。情報処理装置１乃至５（２００）からデータ入出力要求を受信した場合におけるストレージ制御装置１００の基本的な動作は、第一実施例のストレージシステムにおける場合と基本的に同様である。

【００６７】

管理コンピュータ（ＳＶＰ）１６０は、ストレージシステム６００を保守・管理するためのコンピュータである。管理コンピュータ１６０は、ストレージ制御装置１００の内部に設けられているＬＡＮである内部ＬＡＮ５６０を介してチャネル制御部１乃至８（１１０）及びディスク制御部１乃至４（１４０）と接続している。

【００６８】

管理コンピュータ１６０を操作することにより、例えば、ディスクドライブ３００の設定や、論理ボリュームの設定、チャネル制御部１乃至８（１１０）において実行されるマイクロプログラムのインストール等を行うことができる。ここで、ディスクドライブ３００の設定としては、例えば、ディスクドライブ３００の増設や減設、ＲＡＩＤ構成の変更（例えばＲＡＩＤ１からＲＡＩＤ５への変更

等）等を行うことができる。さらに管理コンピュータ１６０からは、ストレージシステム６００の動作状態の確認や故障部位の特定、チャネル制御部１乃至４（ＣＨＮ）１１０で実行されるオペレーティングシステムやアプリケーションプログラムのインストール等の作業を行うこともできる。

【００６９】

管理コンピュータ１６０は、ストレージ制御装置１００に内蔵されている形態とすることもできるし、外付けされている形態とすることもできる。管理コンピュータ１６０は、ストレージ制御装置１００及びディスクドライブ３００の保守・管理を専用に行うコンピュータとすることもできるし、汎用のコンピュータに保守・管理機能を持たせたものとすることもできる。

【００７０】

チャネル制御部１乃至４（ＣＨＮ）１１０で動作しているオペレーティングシステム７０１上では、上述したクラスタソフトウェア１６０以外にも、ＲＡＩＤマネージャ、ボリュームマネージャ、ファイルシステムプログラム、ＮＦＳ（Network File System）等の様々なソフトウェアが動作している。このうち、ＲＡＩＤマネージャは、情報処理装置１乃至５（２００）のユーザやオペレータ等が、ディスク制御装置１４０に対してパラメータの設定や制御を行うためのソフトウェアである。設定されるパラメータの種類としては、例えば、ＲＡＩＤグループを構成するディスクドライブ３００（物理ディスク）を定義（ＲＡＩＤグループの構成情報、ストライプサイズの指定など）するためのパラメータ、ＲＡＩＤレベル（例えば０，１，５）を設定するためのパラメータ等がある。ボリュームマネージャは、ＲＡＩＤ制御部７４０によって提供されるＬＵをさらに仮想化した仮想化論理ボリュームをファイルシステムプログラム７０３に提供する。１つの仮想化論理ボリュームは１以上の論理ボリュームによって構成される。ファイルシステムプログラムは、ネットワーク制御部７０２が受信したファイルアクセス要求に指定されているファイル名とそのファイル名が格納されている仮想化論理ボリューム上のアドレスとの対応づけを管理する。例えば、ファイルシステムプログラム７０３はファイルアクセス要求に指定されているファイル名に対応する仮想化論理ボリューム上のアドレスを特定する。また、ＮＦＳ（Network File

System）７１１は、ＮＦＳ７１１が動作するＵＮＩＸ（登録商標）系の情報処理装置２００からのファイルアクセス要求を受け付ける。

【００７１】

以上に説明した発明の実施の形態は、本発明の理解を容易にするためのものであり、本発明を限定するものではない。本発明は、その趣旨を逸脱することなく、変更、改良され得ると共に、本発明にはその等価物が含まれることは勿論である。

【００７２】

【発明の効果】

本発明によれば、ストレージ制御装置の信頼性や可用性を向上させることができる。

【図面の簡単な説明】

【図１】　本発明の一実施例によるストレージシステムの概略構成を示す図である。

【図２】　本発明の一実施例による情報処理装置のハードウェア構成を示す図である。

【図３】　本発明の一実施例によるチャネル制御部１（ＣＨＮ）のハードウェア構成を示す図である。

【図４】　本発明の一実施例によるチャネル制御部２（ＣＨＦ）のハードウェア構成を示す図である。

【図５】　本発明の一実施例によるディスク制御部のハードウェア構成を示す図である。

【図６】　本発明の一実施例による、データ入出力要求に応じて実行される処理が制限される際の処理を説明するフローチャートを示す図である。

【図７】　本発明の一実施例による情報処理装置－ＷＷＮ対応管理テーブルを示す図である。

【図８】　本発明の一実施例による、データ入出力要求に応じて実行される処理が制限される際の処理の他の例を説明するフローチャートを示す図である。

【図９】　本発明の一実施例によるストレージシステムの他の形態を示す図

である。

【符号の説明】

１００　ストレージ制御装置

１０５　バスライン

１１０　チャネル制御部

１１１　ネットワークインタフェース部

１１２　ＣＰＵ

１１３　メモリ

１１７　通信ポート

１１９　Ｉ／Ｏプロセッサ

１２０　共有メモリ

１３０　キャッシュメモリ

１５０　接続部

２００　情報処理装置

２１９１　通信ポート

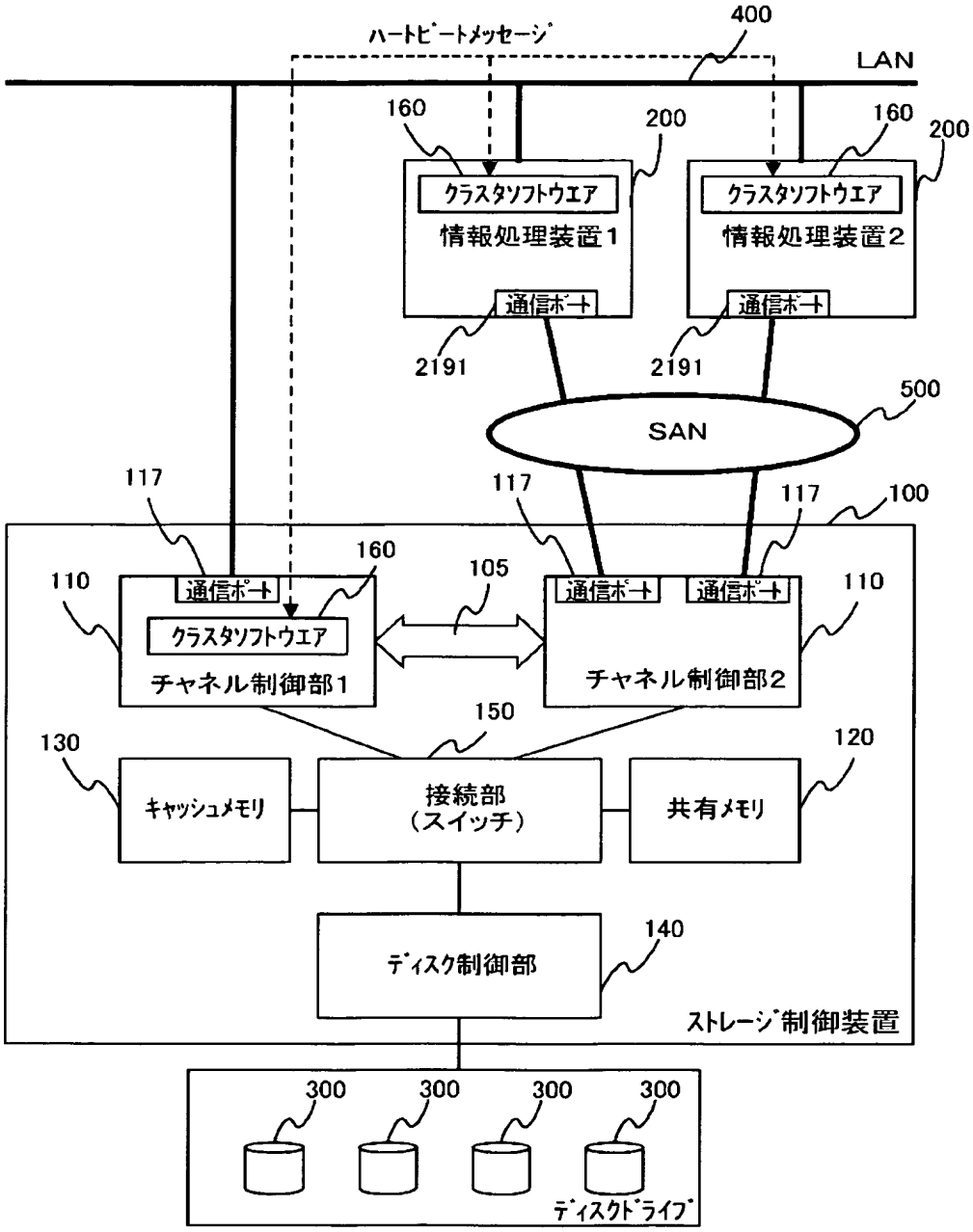
３００　ディスクドライブ

４００　ＬＡＮ

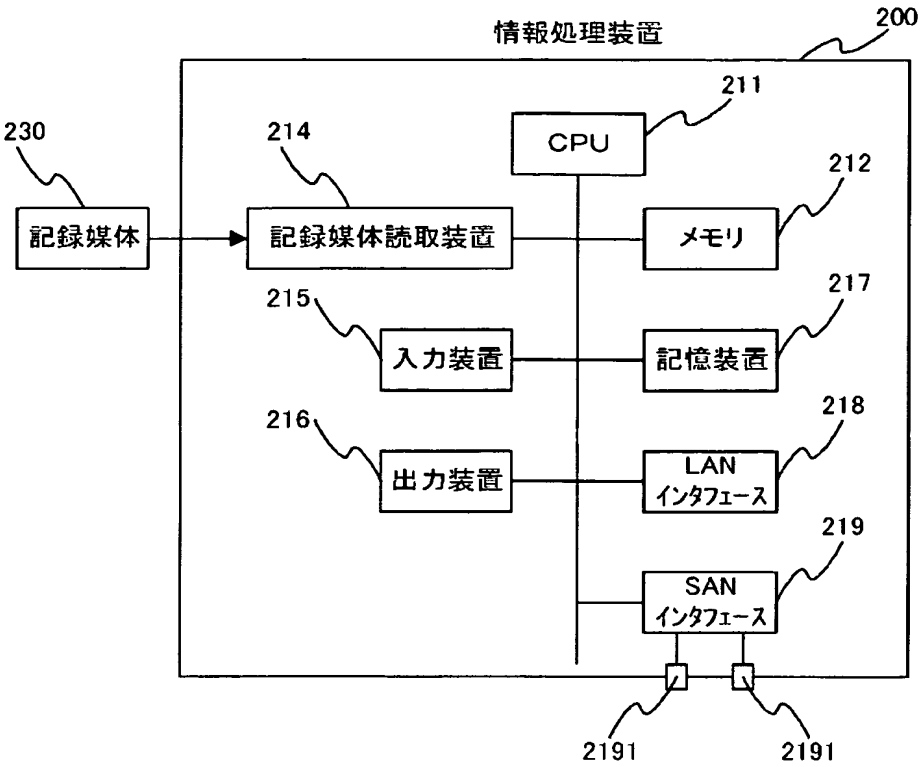
５００　ＳＡＮ

【書類名】　図面

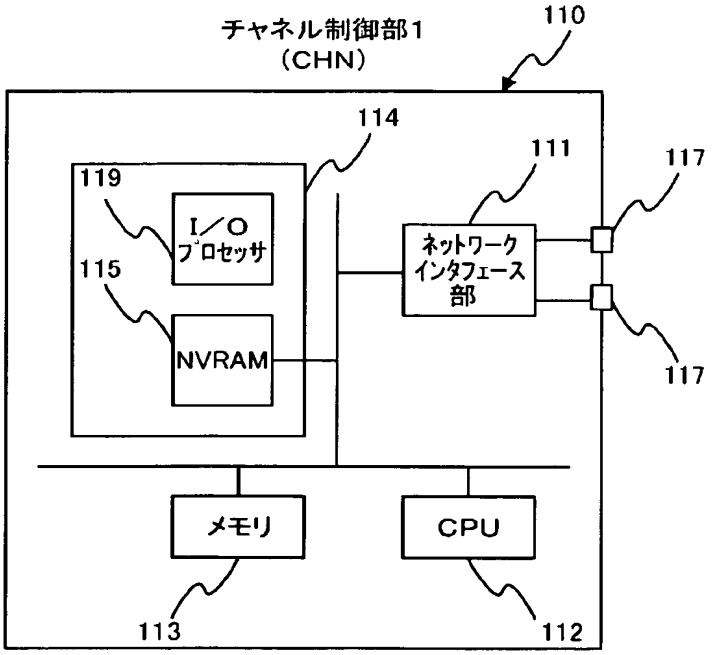
【図１】


ハートビートメッセージ
400
LAN
160
200
160
200
クラスタソフトウエア
情報処理装置１
クラスタソフトウエア
情報処理装置２
通信ポート
通信ポート
2191
2191
500
SAN
117
117
117
100
110
160
105
通信ポート
通信ポート
通信ポート
110
クラスタソフトウエア
チャネル制御部１
チャネル制御部２
150
130
120
キャッシュメモリ
接続部
（スイッチ）
共有メモリ
140
ディスク制御部
ストレージ制御装置
300
300
300
300
ディスクドライブ

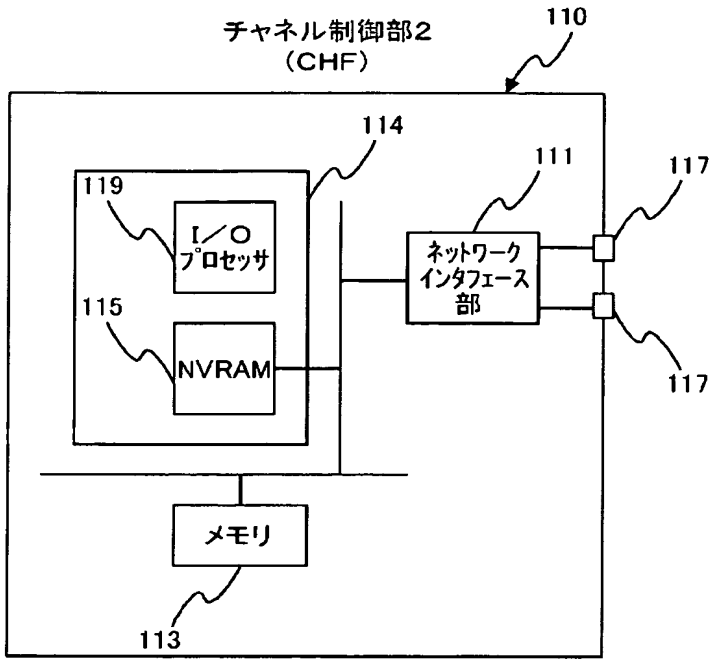
【図２】


情報処理装置
200
211
CPU
230
214
記録媒体
記録媒体読取装置
212
メモリ
215
217
入力装置
記憶装置
216
218
出力装置
LAN
インタフェース
219
SAN
インタフェース
2191
2191

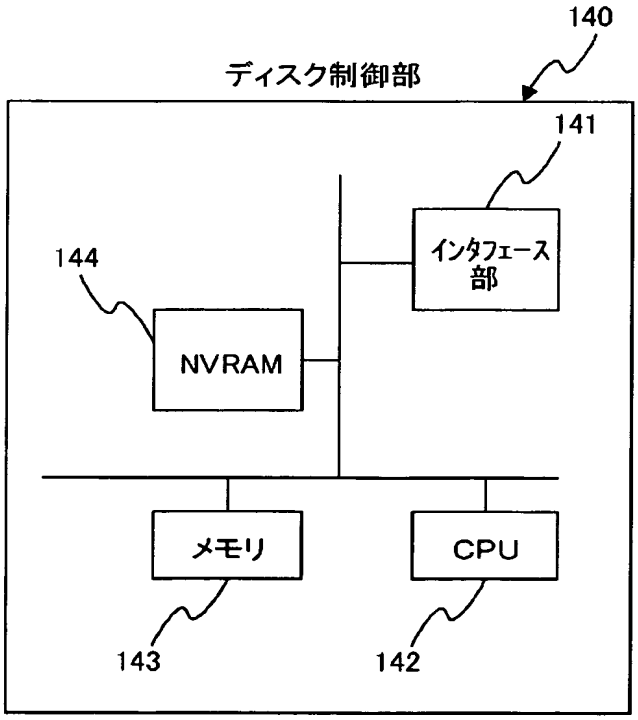
【図３】


チャネル制御部1
(CHN)
110
114
111
117
119
I/O
プロセッサ
115
ネットワーク
インタフェース
部
NVRAM
117
メモリ
CPU
113
112

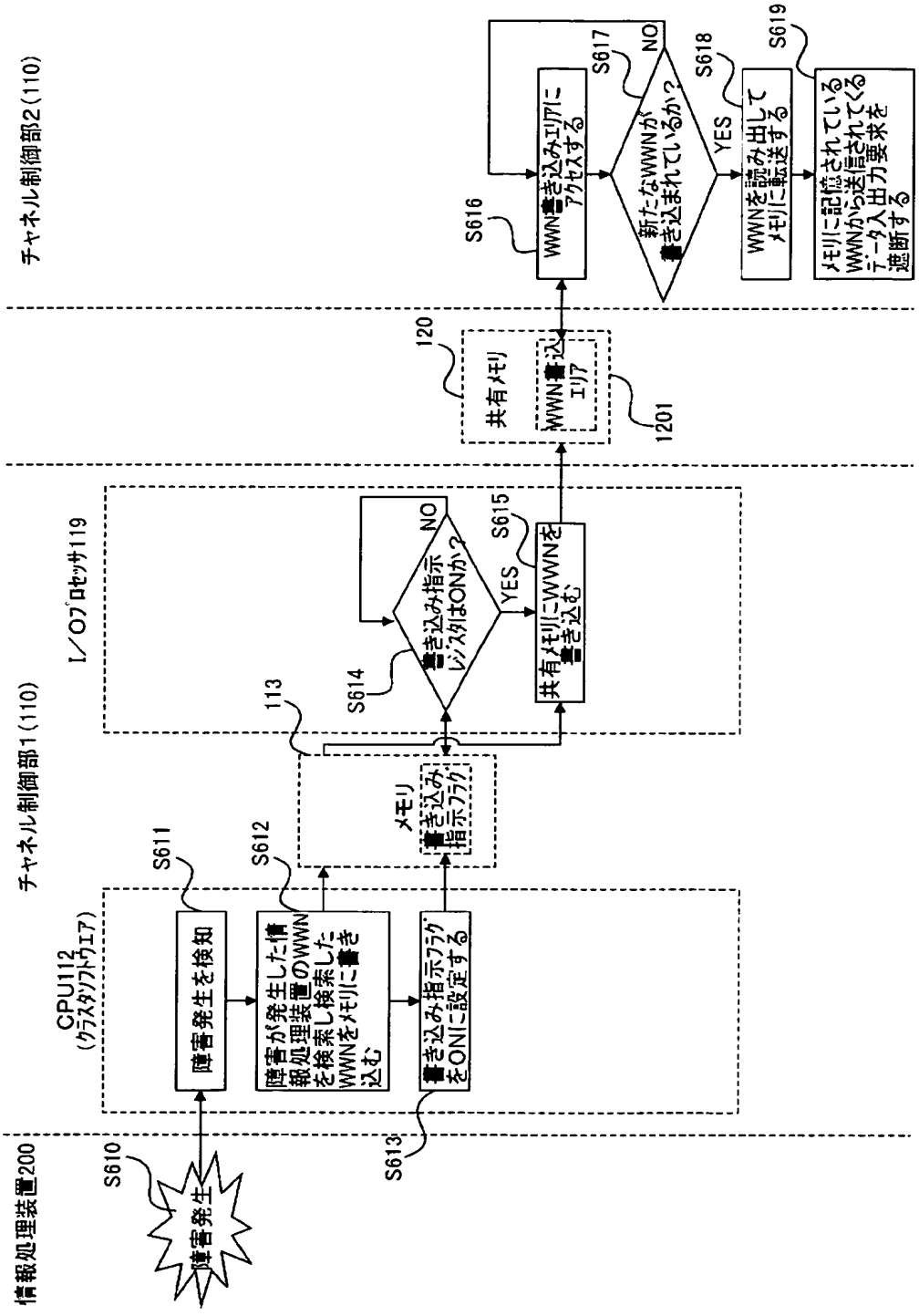
【図４】


チャネル制御部2
(CHF)
110
114
111
117
119
I/O
プロセッサ
ネットワーク
インタフェース
部
115
NVRAM
117
メモリ
113

【図５】


140
ディスク制御部
141
インタフェース
部
144
NVRAM
メモリ
CPU
143
142

【図６】


情報処理装置200
障害発生
S610
チャネル制御部1(110)
CPU112
(ファームウェア)
障害発生を検知
S611
障害が発生した情報処理装置のWWNを検索し検索したWWNをメモリに書き込む
S612
書き込み指示フラグをONに設定する
S613
メモリ
書き込み指示フラグ
113
L/Oプロセッサ119
書き込み指示レジスタはONか？
S614
NO
YES
S615
共有メモリにWWNを書き込む
120
共有メモリ
WWN書込エリア
1201
チャネル制御部2(110)
WWN書き込みエリアにアクセスする
S616
新たなWWNが書き込まれているか？
S617
NO
YES
S618
WWNを読み出してメモリに転送する
メモリに記憶されているWWNから送信されてくるデータ入出力要求を遮断する
S619

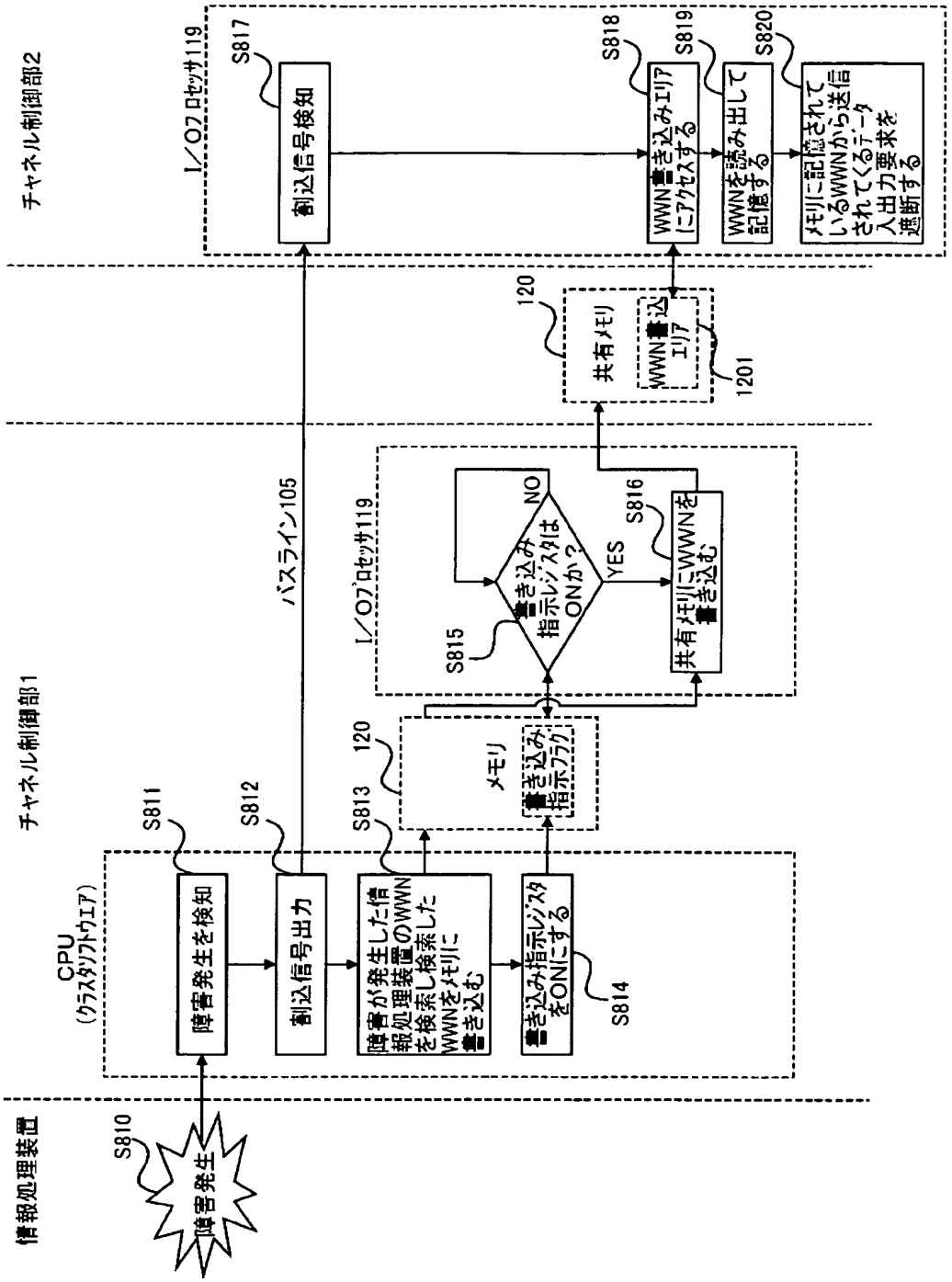
【図７】

700

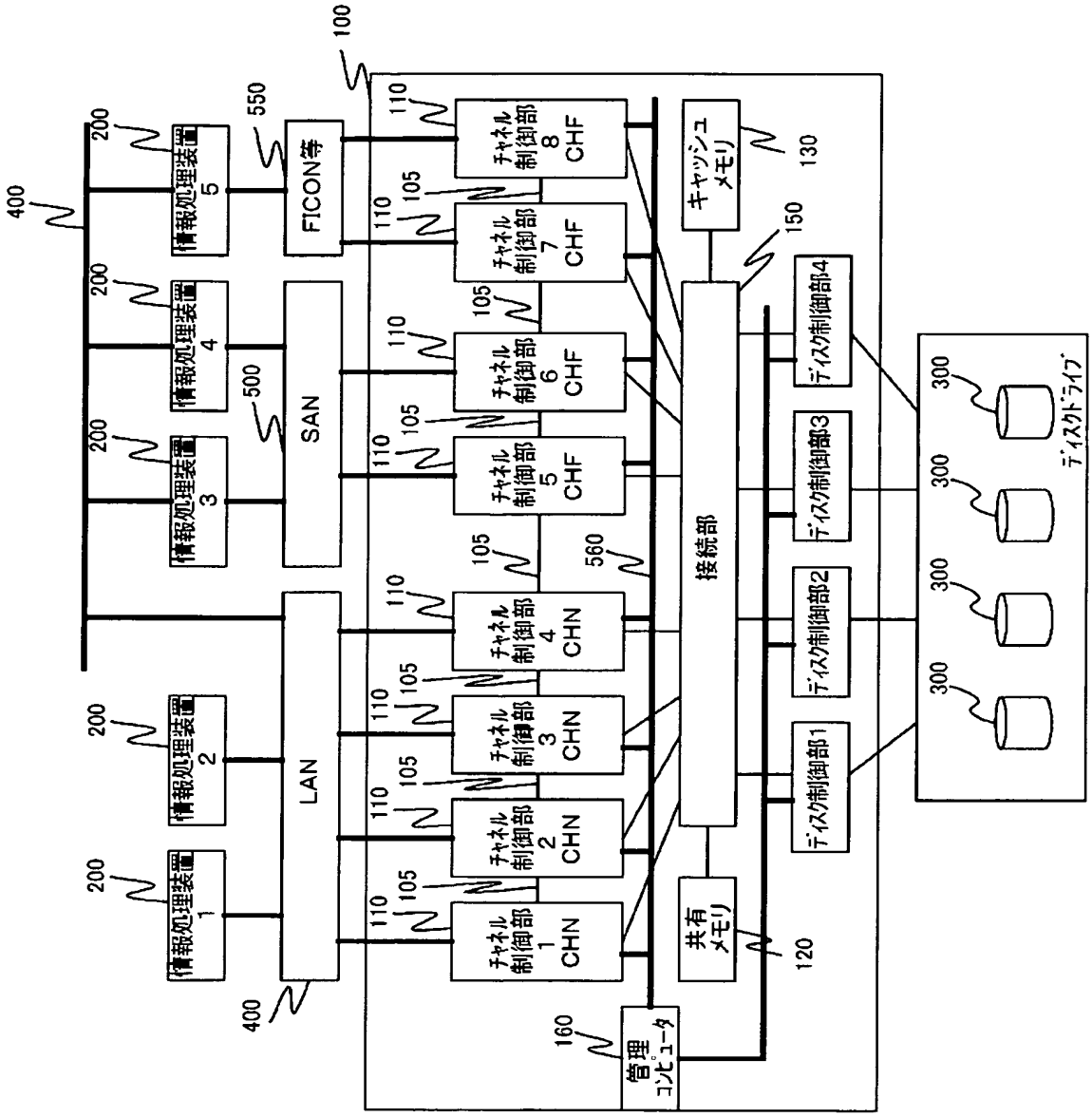
情報処理装置-WWN対応管理テーブル

情報処理装置のIPアドレス	情報処理装置の通信ポートに付与されているWWN
192.168.0.1	10:00:00:00:00:00:00:01
192.168.0.2	10:00:00:00:00:00:00:02
192.168.0.3	12:00:00:00:00:00:00:01
192.168.0.4	12:00:00:00:00:00:00:02
・ ・ ・	・ ・ ・

【図８】


情報処理装置
チャネル制御部1
チャネル制御部2
CPU
(クラスタソフトウエア)
障害発生
障害発生を検知
創込信号出力
障害が発生した情報処理装置のWWNを検索し検索したWWNをメモリに書き込む
書き込み指示レジスタをONにする
メモリ
書き込み指示フラグ
120
パスライン105
I/Oプロセッサ119
書き込み指示レジスタはONか？
NO
YES
共有メモリにWWNを書き込む
共有メモリ
WWN書込エリア
1201
創込信号検知
WWN書き込みエリアにアクセスする
WWNを読み出して記憶する
メモリに記憶されているWWNから送信されてくるデータ入出力要求を遮断する
S810
S811
S812
S813
S814
S815
S816
S817
S818
S819
S820

【図９】


100
400
200
情報処理装置 1
情報処理装置 2
情報処理装置 3
情報処理装置 4
情報処理装置 5
550
FICON等
500
SAN
LAN
400
110
105
チャネル制御部 1 CHN
チャネル制御部 2 CHN
チャネル制御部 3 CHN
チャネル制御部 4 CHN
560
チャネル制御部 5 CHF
チャネル制御部 6 CHF
チャネル制御部 7 CHF
チャネル制御部 8 CHF
接続部
キャッシュメモリ
130
150
共有メモリ
120
160
管理コンピュータ
ディスク制御部1
ディスク制御部2
ディスク制御部3
ディスク制御部4
300
ディスクドライブ

【書類名】　　要約書

【要約】

【解決手段】　情報処理装置から送信されるデータ入出力要求を受信して、前記データ入出力要求に応じて記憶デバイスに対するデータの書き込み／読み出しを制御するための制御信号を出力する、互いに通信可能に接続される複数のチャネル制御部と、前記制御信号に応じて記憶デバイスに対するデータの書き込み／読み出しを行うディスク制御部と、を備えるストレージ制御装置の制御方法であって、第１の前記チャネル制御部が、前記情報処理装置と通信することにより前記情報処理装置の動作を監視し、第１の前記チャネル制御部が、前記情報処理装置における障害を検知した場合に、前記情報処理装置から送信されたデータ入出力要求が着信される第２の前記チャネル制御部により前記データ入出力要求に応じて実行される処理を制限するための処理を実行するようにする。

【選択図】　　図１

特願２００３－１５８２７１

出願人履歴情報

識別番号　［０００００５１０８］

１．変更年月日　１９９０年　８月３１日
［変更理由］　新規登録
住　所　東京都千代田区神田駿河台４丁目６番地
氏　名　株式会社日立製作所